

CLAIMS

1. An array comprising a plurality of nucleic acid members, each member having a unique position and stably associated with a solid substrate, wherein each nucleic acid member comprises a non-coding sequence present in a 3'-end of an RNA transcript, and wherein each of said nucleic acid members is less than 600 nucleotides.
2. An array comprising a plurality of nucleic acid members, each member having a unique position and stably associated with a solid substrate, wherein each nucleic acid member comprises a non-coding sequence present in a 5' end of an RNA transcript, and wherein each of said nucleic acid members is less than 1000 nucleotides.
3. The array of claim 1 or 2, wherein said noncoding sequence is at least 20 nucleotides in length.
4. ~~The array of claim 1 or 2, wherein each said nucleic acid member comprises substantially noncoding sequences.~~
5. The array of claim 1, wherein said nucleic acid members comprise human sequences.
6. The array of claim 5, wherein at least one position on said array comprises a control position comprising a substance selected from the group consisting of: a buffer, a cDNA encoded by a housekeeping gene, a plant gene sequence, and a vector sequence.
7. The array of claim 1 or 2, wherein said array comprises from 1000 to 10,000 positions.
8. The array of claim 1 or 2, wherein at least 2% of the nucleic acid members on the array comprise sequences which are not included within a public database.

9. The array of claim 5, wherein said nucleic acid members comprise sequences expressed in at least two different tissues.
10. The array of claim 1 or 2, wherein said nucleic acid members comprise sequences expressed in a healthy tissue.
11. The array of claim 1 or 2, wherein said nucleic acid members comprise sequences expressed in a diseased tissue.
12. The array of claim 1 or 2, wherein said nucleic acid members comprise sequences expressed in a tissue which has been exposed to a drug.
13. The array of claim 1 or 2 wherein said nucleic acid members do not comprise repeat sequences.
14. A composition comprising a set of nucleic acid members, each nucleic acid member comprising a non-coding sequence present in a 3'-end of an RNA transcript, and wherein each of said nucleic acid members is less than 600 nucleotides.
15. A composition comprising a plurality of nucleic acid members, each nucleic acid member comprising a non-coding sequence present in a 5'-end of an RNA transcript, and wherein each of said nucleic acid members is less than 1000 nucleotides.
16. The composition of claim 14 or 15, wherein said noncoding sequence is at least 20 nucleotides in length.
17. The composition of claim 14 or 15, wherein said nucleic acid members comprise substantially noncoding sequences.
18. The composition of claim 14 or 15, wherein said nucleic acid members comprise human sequences.

19. The composition of claim 18, further including a control sequence selected from the group consisting of: a cDNA encoded by a housekeeping gene, a plant gene sequence, and a vector sequence.
20. The composition of claim 14 or 15, comprising at least 1000 nucleic acid members which are not identical in sequence.
21. The composition of claim 14 or 15, wherein at least 2% of the nucleic acid members of the composition comprise sequences which are not included within a public database.
22. The composition of claim 21, wherein said nucleic acid members comprise sequences expressed in at least two different tissues.
23. The composition of claim 14 or 15, wherein said nucleic acid members comprise sequences expressed in a healthy tissue.
24. The composition of claim 14 or 15, wherein said nucleic acid members comprise sequences expressed in a diseased tissue.
25. The composition of claim 14 or 15, wherein said nucleic acid members comprise sequences expressed in a tissue which has been exposed to a drug.
26. The composition of claim 14 or 15 wherein said nucleic acid members do not comprise repeat sequences.

32. The method of claim 27 or 28, wherein said nucleic acid members are each assigned an identifier based on whether said nucleic acid member is a known or unknown sequence.
33. The method of any of claims 27 or 28, wherein data relating to the hybridization of a target sequence to a nucleic acid member is stored in a database.
34. A method for producing a composition comprising a plurality of nucleic acid members, each nucleic acid member comprising a noncoding sequence present in a 3'-end or a 5' end of an RNA transcript, said method comprising the steps of:
 - a) selecting a cDNA molecule from a population of cDNAs
 - b) sequencing at least a portion of said 3'-end or said 5' end of said cDNA to obtain sequence information about said cDNA;
 - c) providing a primer complementary to a subsequence within said portion of said 3'-end, and amplifying said cDNA, thereby producing a nucleic acid member;
 - d) repeating steps (a) through (c) at least once; and
 - e) selecting a cDNA which comprises substantially noncoding sequences and excludes repeat elements.
35. The method of claim 34, further comprising the step of comparing said sequence information with sequence information in a public database.
36. The method of claim 35, further comprising assigning said cDNA an identifier and classifying said cDNA as known, if said sequence of said cDNA is at least 95% identical over 100 nucleotides to a sequence in said database, after maximally aligning said sequence of said cDNA to said sequence in said database; and classify said cDNA as unknown if said sequence is less than 95% identical;
37. The method of claim 34, wherein said population of cDNA molecules is a cDNA library.

38. The method of claim 34, wherein said population of cDNA molecules is reverse transcribed from a population of RNA molecules.

39. The method of ~~claim 34~~, wherein at least two of said nucleic acid members is stably associated with a substrate, and each nucleic acid member is placed at a different position on the substrate.

40. The method of claim 34 or 38, wherein said population of cDNA molecules comprise human sequences.

41. The method according to claim 36, wherein at least 2% of said sequences is unknown.

42. The method according to ~~claim 38~~, wherein said sequences are further classified according to whether or not a polyadenylation signal is present.